Introduction

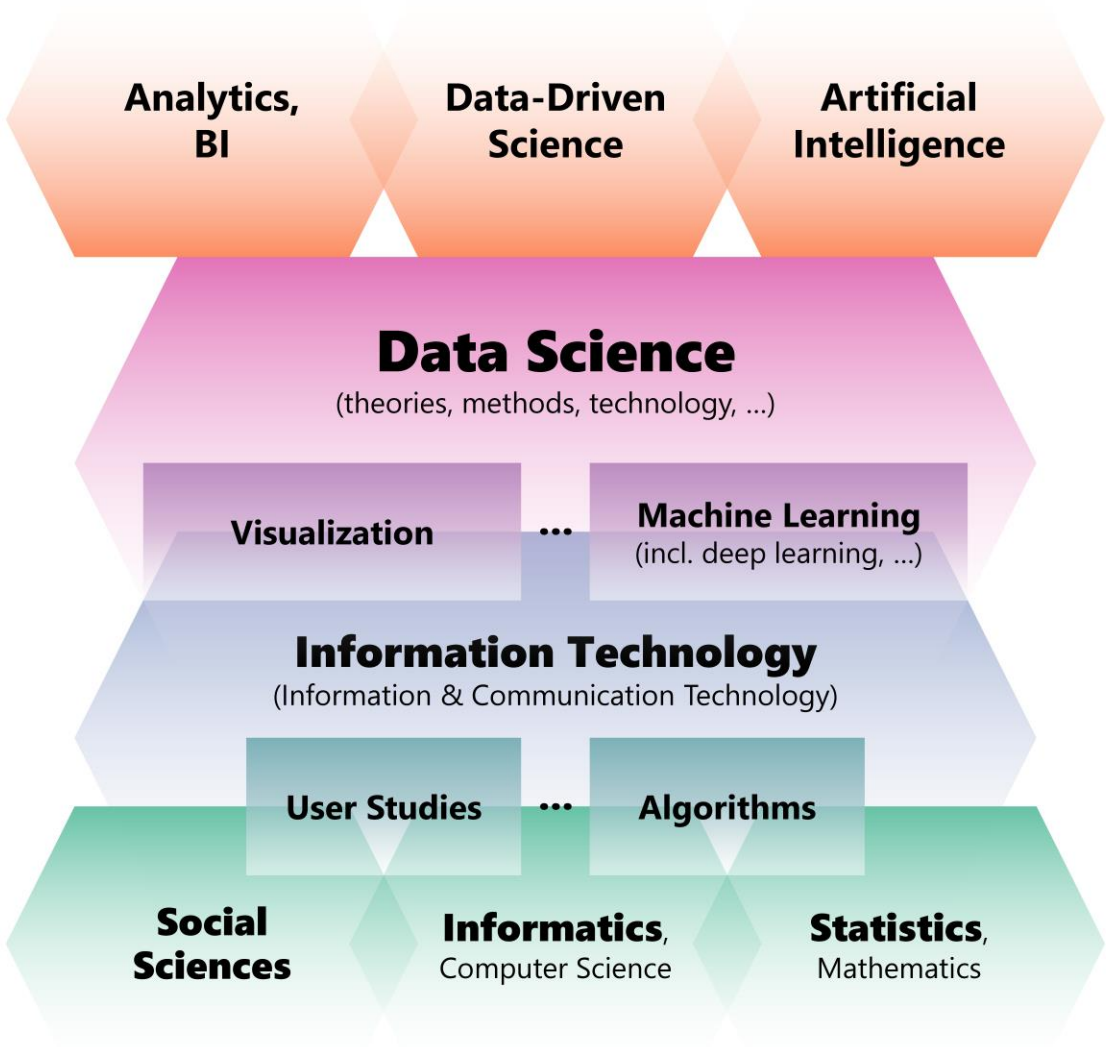
Course Instructor:

Rashedul Alam Shakil

Automation Programmer at
Siemens, Germany.
Founder, Study Mart (YouTube)
Founder, aiQuest Intelligence
Master in Data Science at FAU Erlangen

What, Why, How?

Data Science



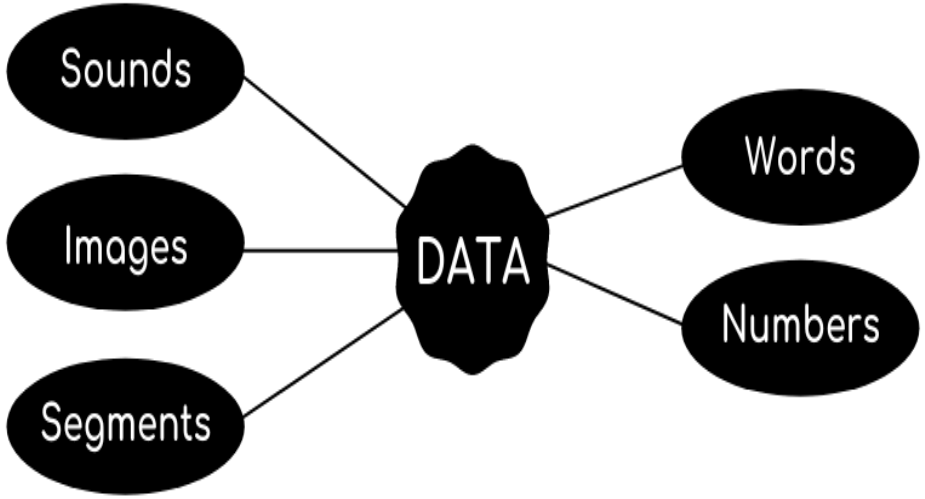
Let’s talk about Data, Information, and Database

What is Data?

Definition

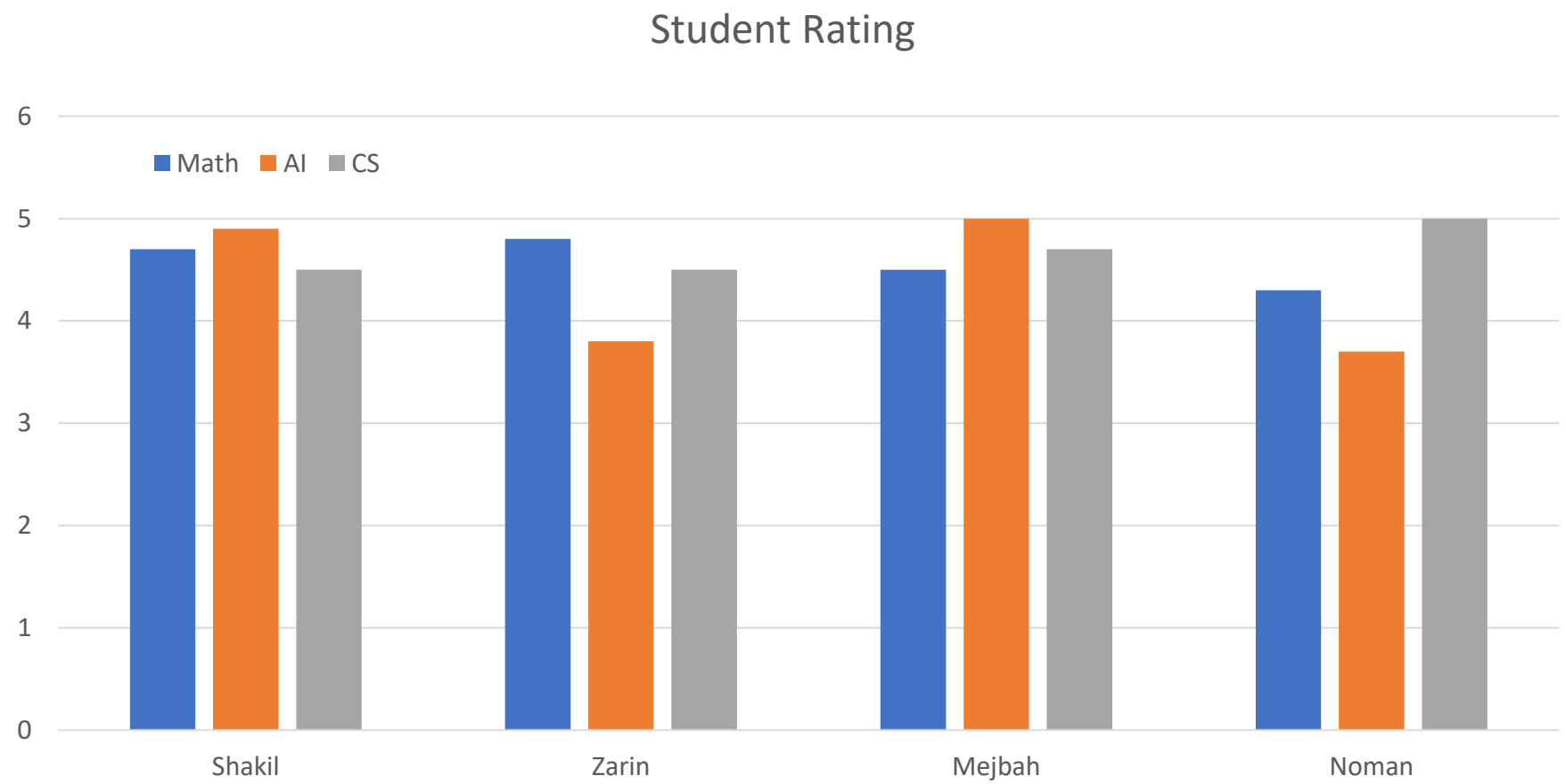
Data is defined as a collection of organized or unorganized facts, concepts, or instructions in a formalized manner suitable for communication, interpretation, processing by humans, or some automatic means such as computers, ATMs.

The main examples of data are phone numbers, weights, prices, costs, number of items sold, product names, addresses, registration marks, etc.



What is Information?

Definition



What is Database?

Definition

A database is an organized collection of related data or information stored and accessed electronically within a computer system.

For example, MySQL, MongoDB, Oracle Database, PostgreSQL, etc. are all examples of different databases. These modern databases are managed by DBMS. Structured Query Language, or SQL as it is more widely known, is used to operate on the data in a database.

How Much Data is in the World Today?

How Much Data is in the World Today?

1 How much data is generated every minute?

Source: Domo

41,666,667
messages shared
by WhatsApp users

1,388,889
video / voice calls made
by people worldwide

404,444
hours of video streamed
by Netflix users

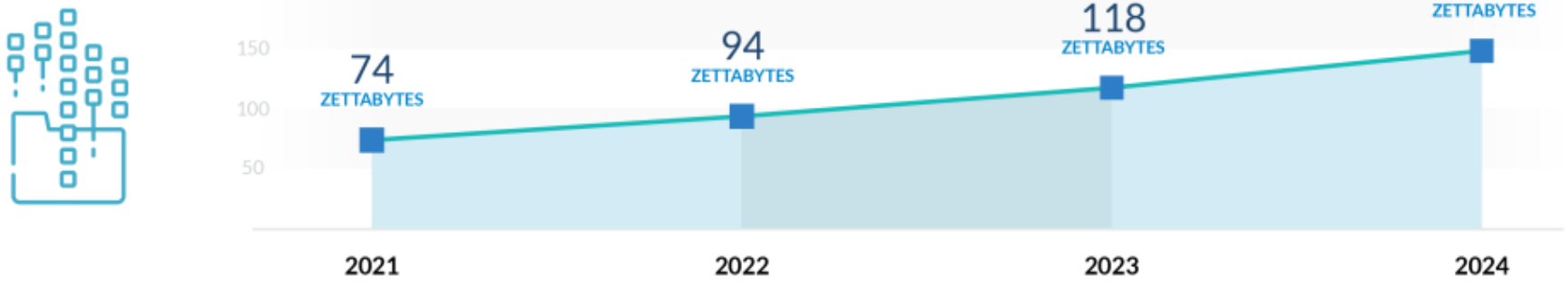
347,222
stories posted by Instagram users

150,000
messages shared by Facebook users

147,000
photos shared by Facebook users

2 Estimated Data Consumption from 2021 to 2024

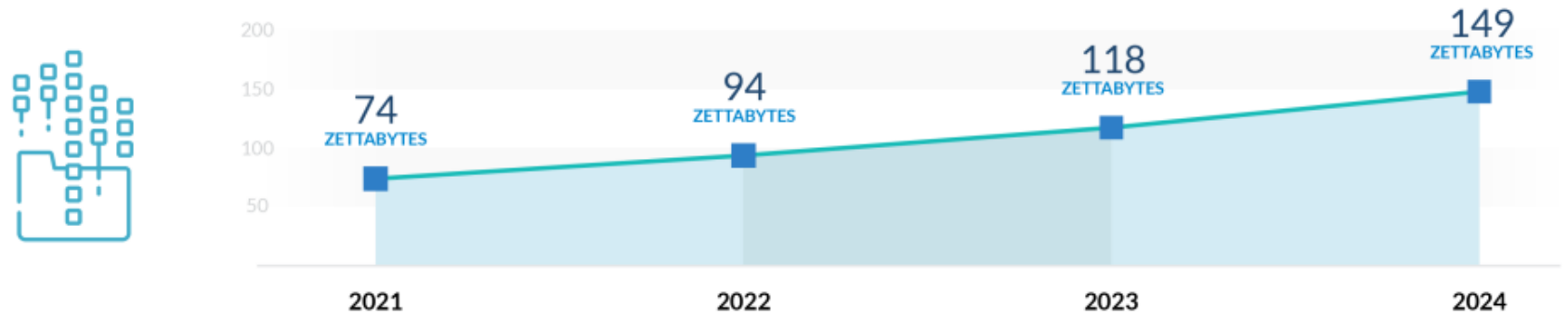
Source: IDC / Statista



How Much Data is in the World Today?

2 Estimated Data Consumption from 2021 to 2024

Source: IDC / Statista



3 Data Growth in 2021

Sources: TechJury, Internet Live Stats, Cisco, PurpleSec

2 TRILLION
searches on Google by the end of 2021

1.134 TRILLION MB
volume of data created every day

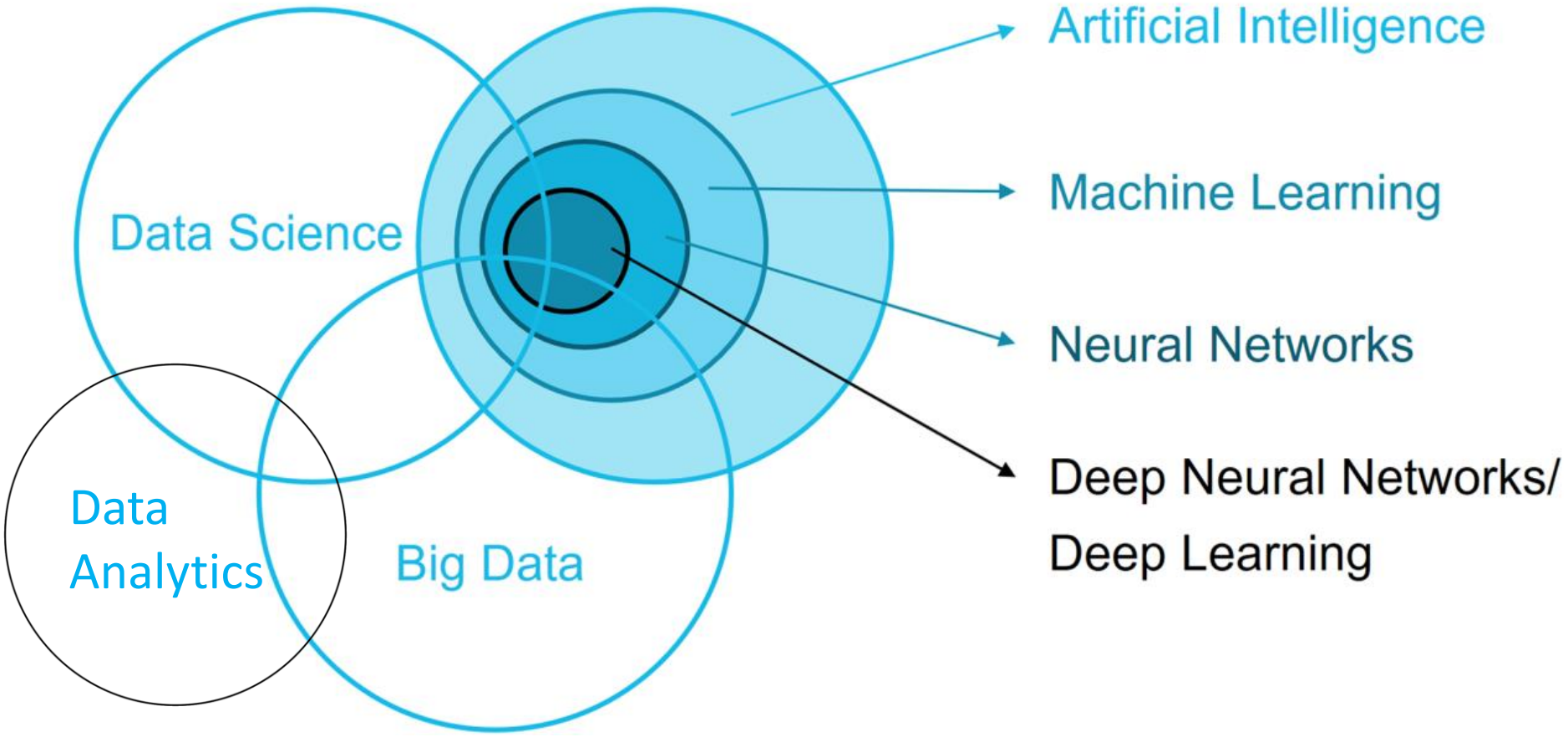
3,026,626
emails sent every second, 67% of which are spam

278,108 PETABYTES
global IP data per month by the end of 2021

230,000
new malware versions created every day

82%
share of video in total global internet traffic at the end of 2021

How to Make Decisions Based on Data?



Data Science:

- Data/Sample
- Mathematics
- Statistics
- Computer Science

Big data: The size of the data is **beyond the ability** of typical database software tools to capture, store, manage, and analyze. This might be data sizes of terabytes, petabytes, or even exabytes. The speed at which the data is created, collected, and processed is extremely high. This requires real-time processing and analysis.

Artificial Intelligence

Core Concept of AI

Artificial Intelligence

AI involves techniques that equip computers to emulate human behavior, enabling them to learn, make decisions, recognize patterns, and solve complex problems in a manner akin to human intelligence.

Machine Learning

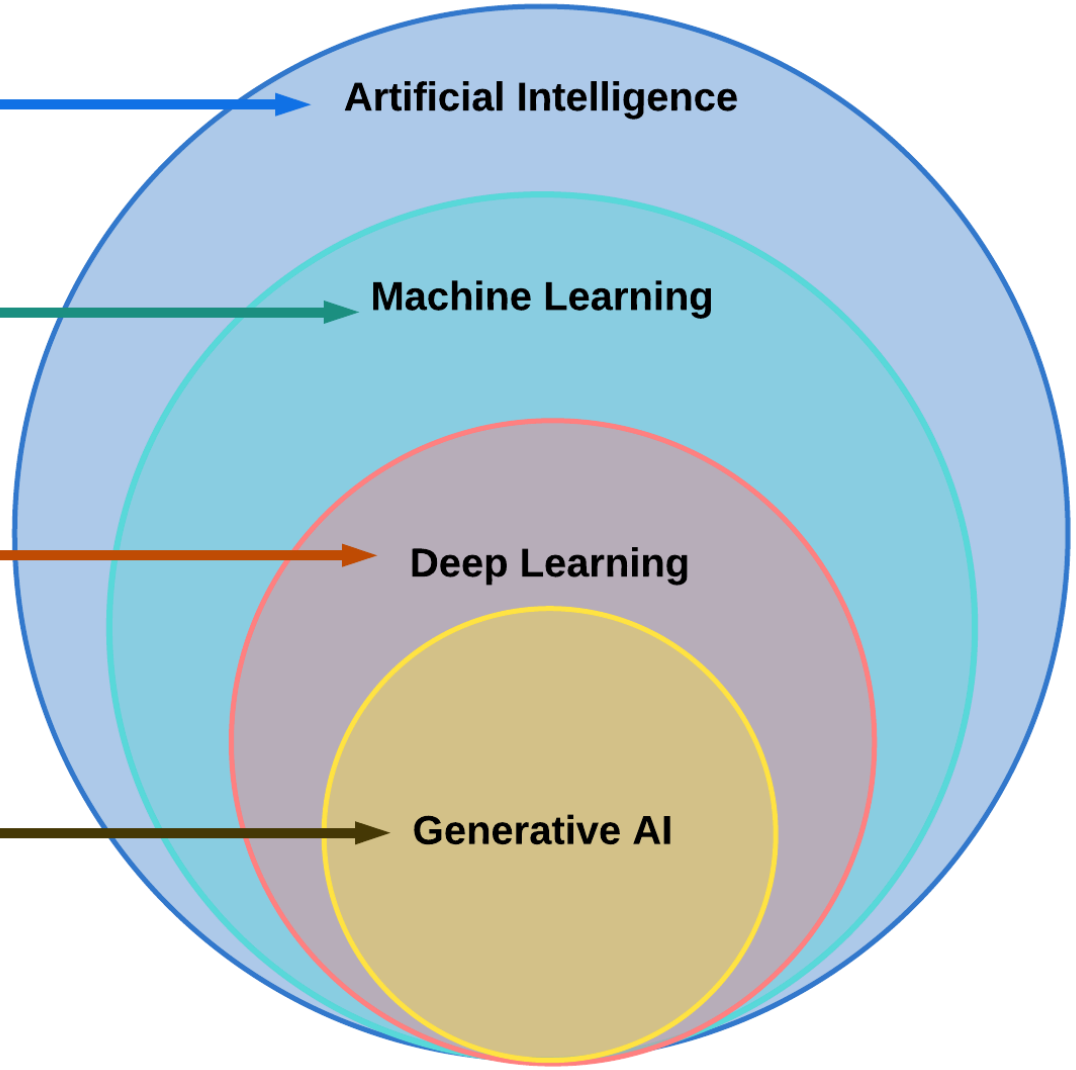
ML is a subset of AI, uses advanced algorithms to detect patterns in large data sets, allowing machines to learn and adapt. ML algorithms use supervised or unsupervised learning methods.

Deep Learning

DL is a subset of ML which uses neural networks for in-depth data processing and analytical tasks. DL leverages multiple layers of artificial neural networks to extract high-level features from raw input data, simulating the way human brains perceive and understand the world.

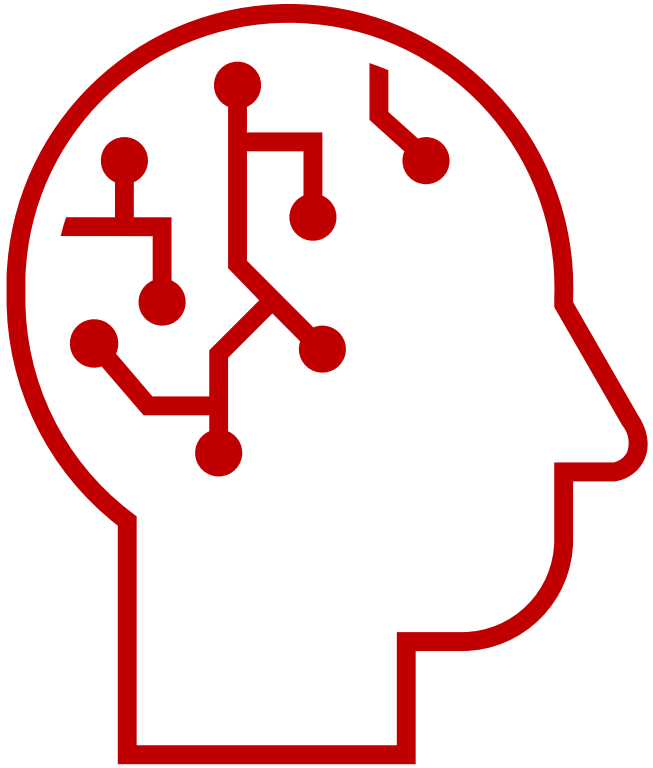
Generative AI

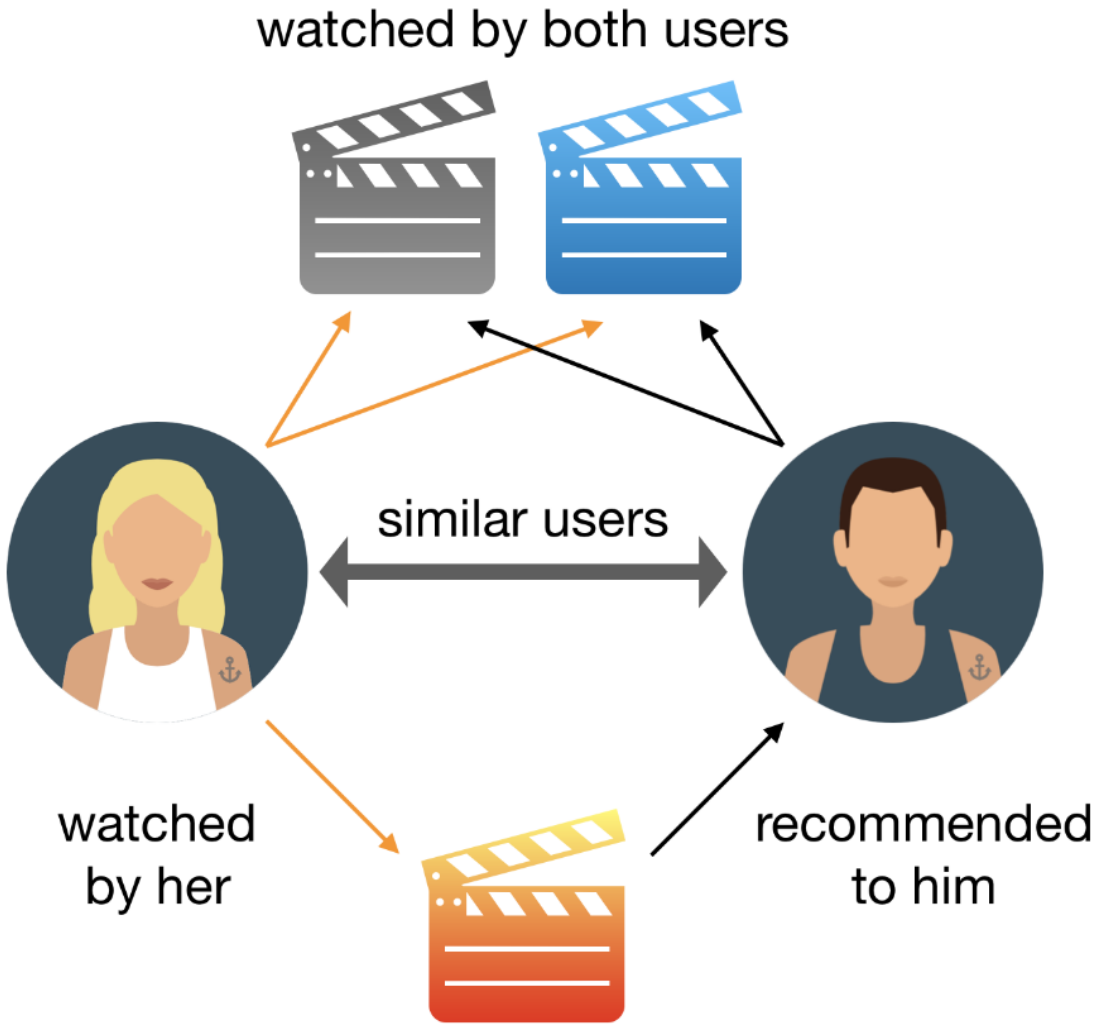
Generative AI is a subset of DL models that generates content like text, images, or code based on provided input. Trained on vast data sets, these models detect patterns and create outputs without explicit instruction, using a mix of supervised and unsupervised learning.



What is AI?

Definition



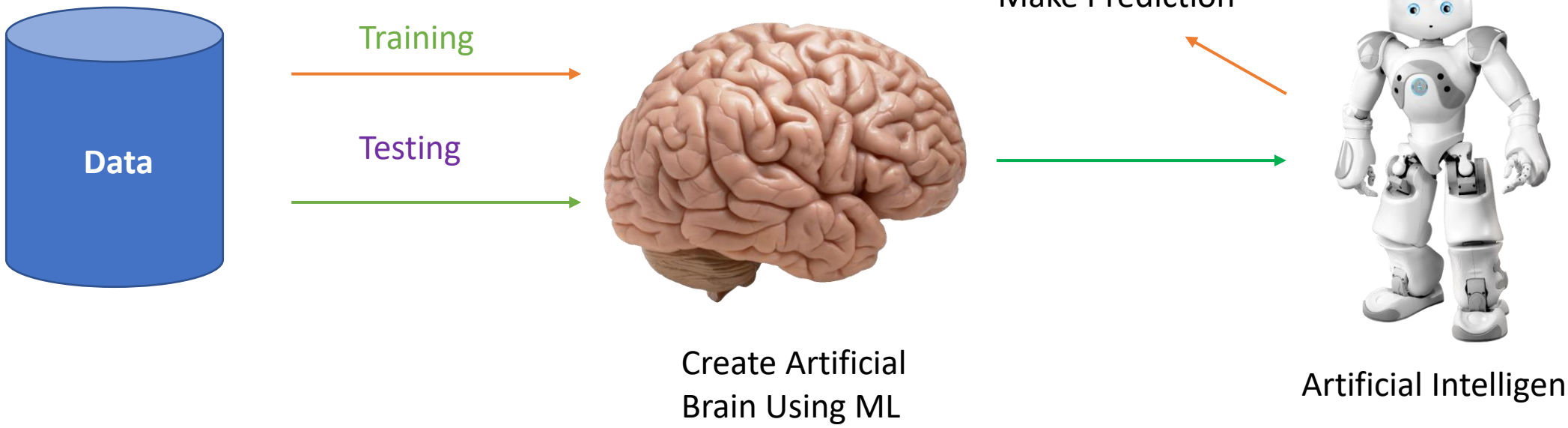


Machine Learning

Definition

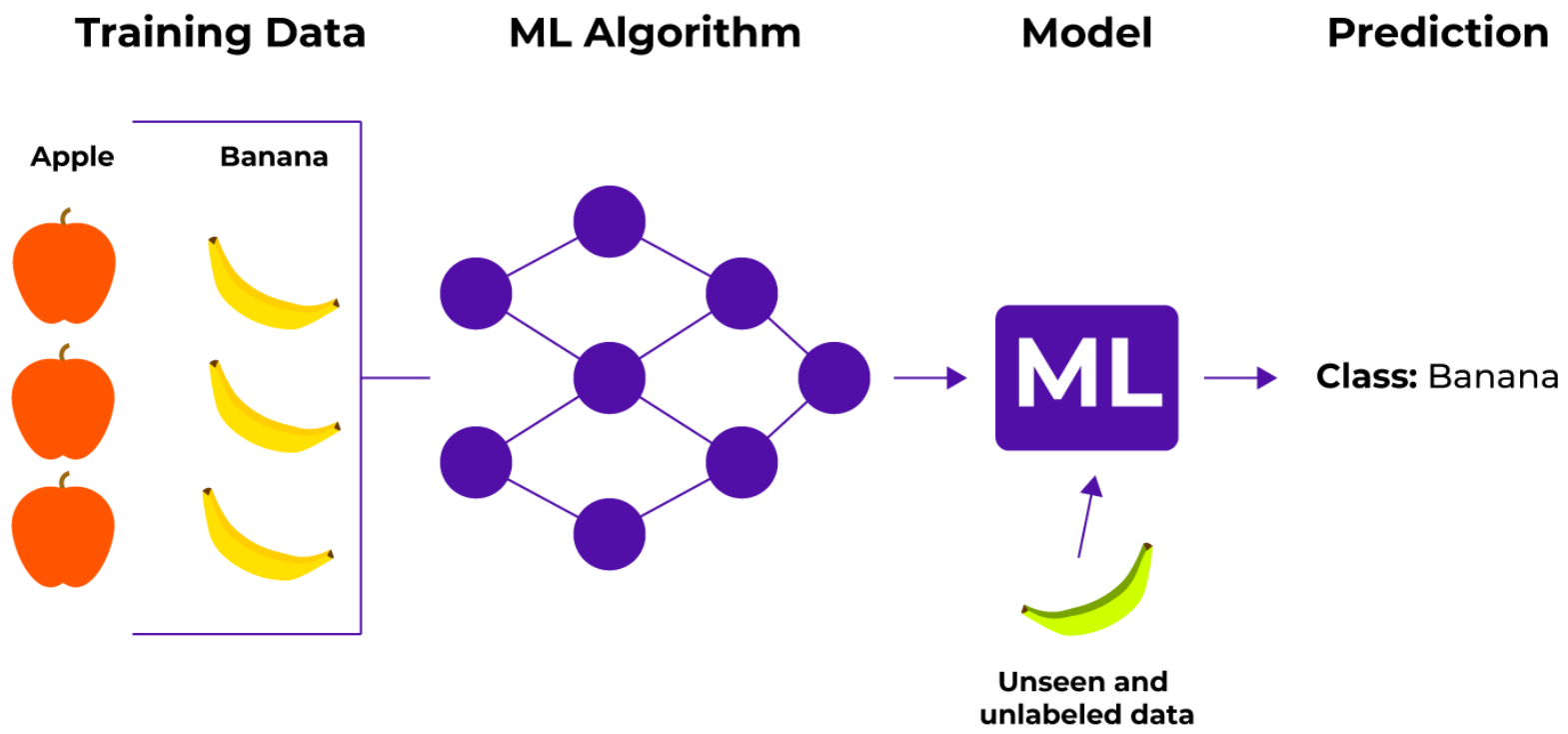
What is Machine Learning?

ML = Machine + Learning



Machine Learning

Example

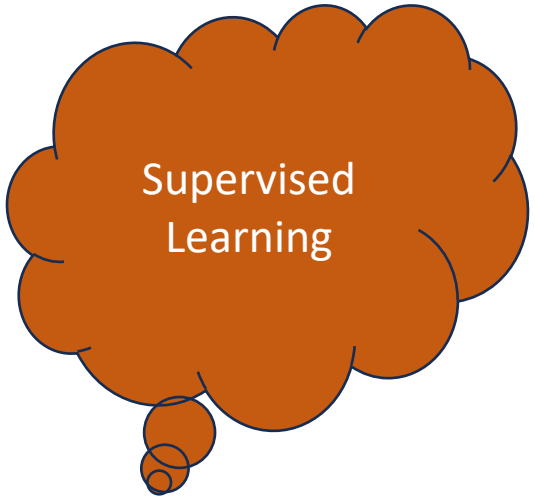


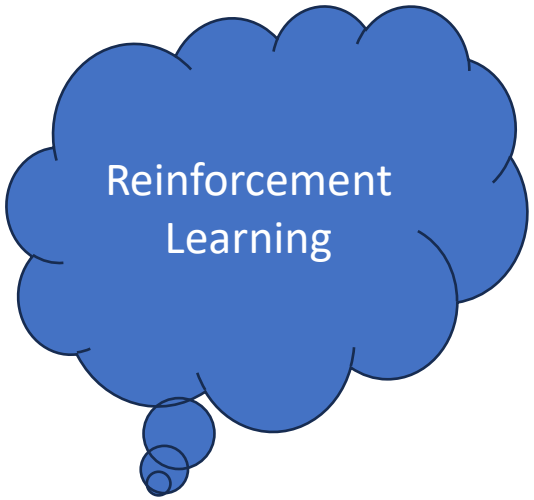
Machine Learning

Working Areas

- Computer Vision
- Natural Language Processing (NLP)
- Speech Recognition
- Reinforcement Learning
- Robotics
- Recommendation Systems
- Healthcare and Bioinformatics
- Finance
- Cybersecurity
- Autonomous Systems

Types of Machine Learning

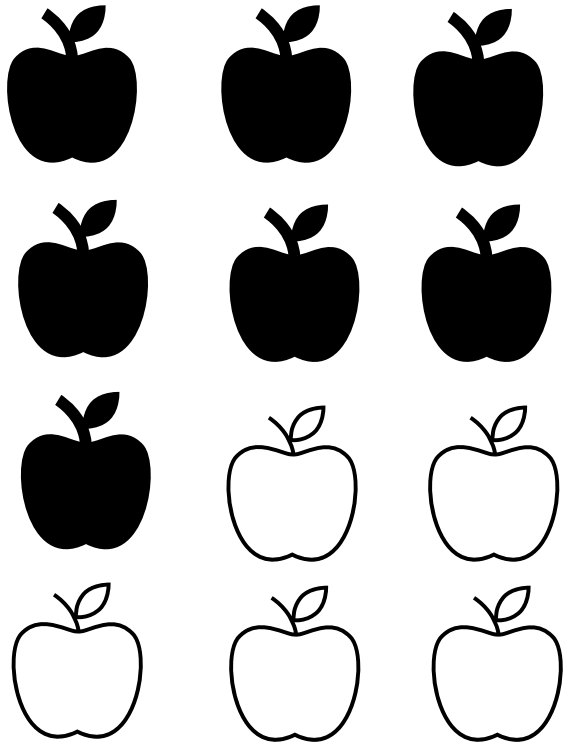




Statistical Machine Learning

Supervised Learning

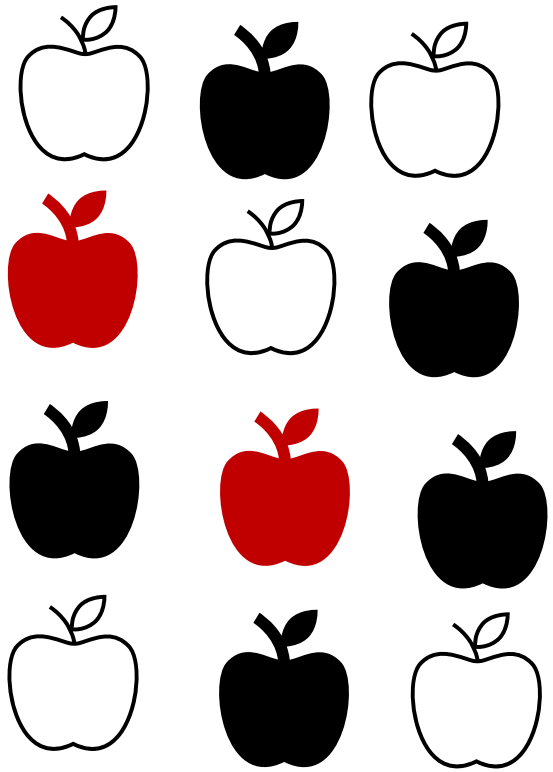
1. Classification Tasks



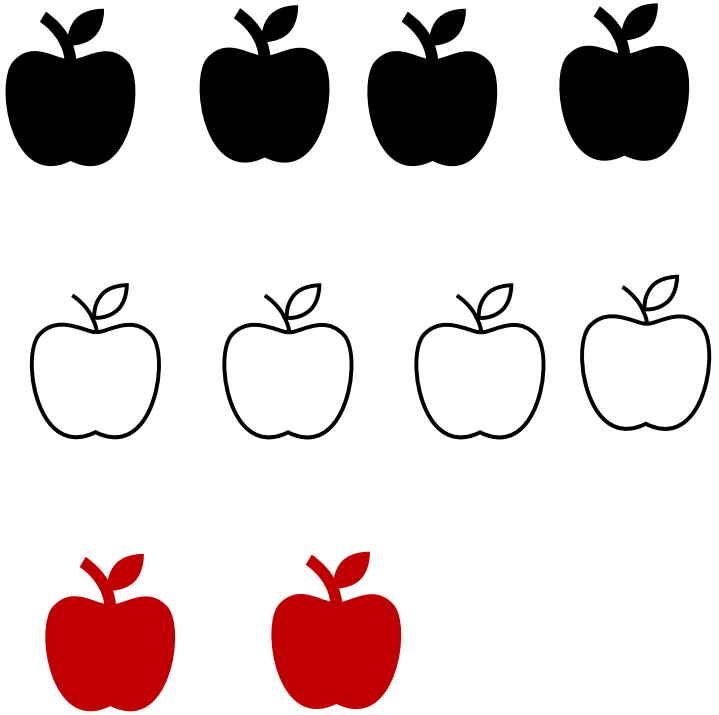
2. Regression Tasks

Statistical Machine Learning

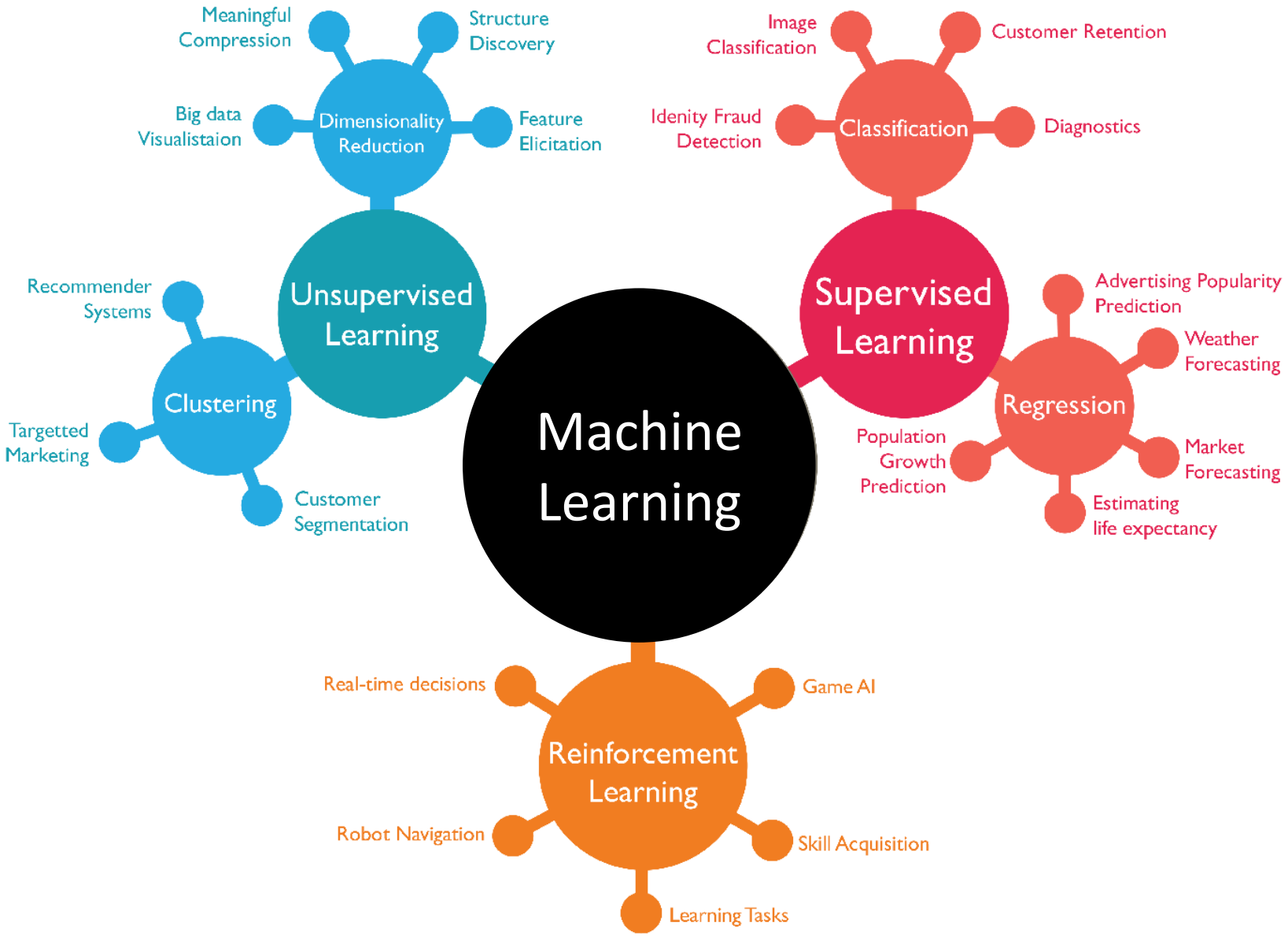
Before Cluster



After Cluster



Statistical Machine Learning



Deep Learning

1. Convolutional Neural Networks (CNNs)
2. Recurrent Neural Networks (RNNs)
3. Long Short-Term Memory (LSTM) Networks
4. Generative Adversarial Networks (GANs)
5. Variational Autoencoders (VAEs)
6. Transformer Networks
7. N-Gram Model

Fig: Training Model

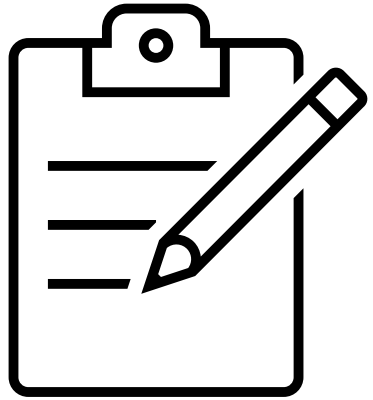
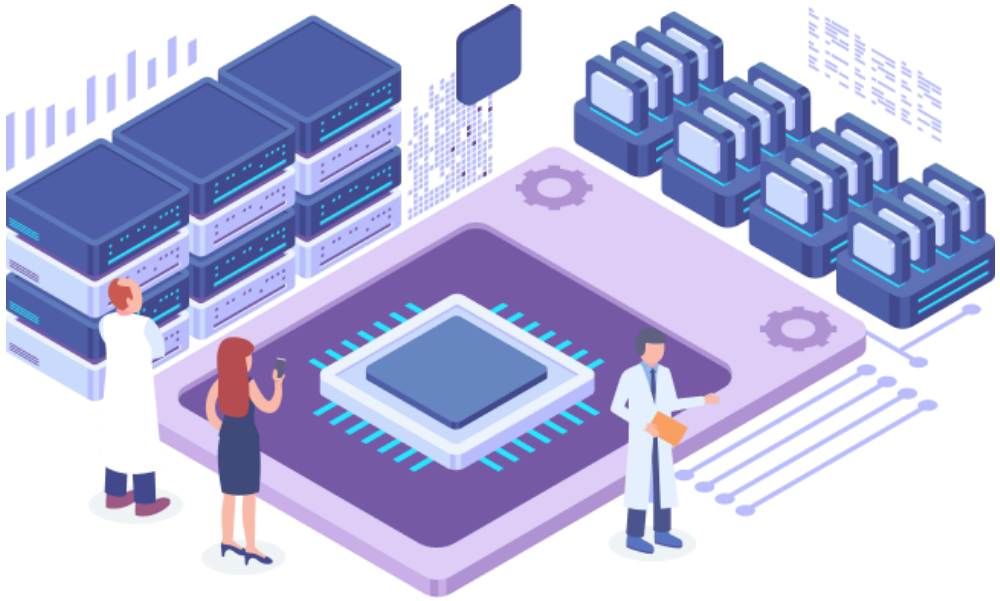


Fig: Generating Answer

Large Language Models: LLMs

Brief Discussion

Large language models (LLMs) are natural language processing computer programs that use artificial neural networks to generate text. Some notable ones are GPT-3, GPT-4, LaMDA (Bard), BLOOM, and LLaMA. LLMs power many applications, such as AI chatbots and AI search engines.



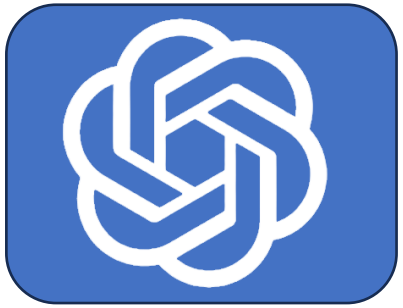
Generative AI Scopes

General Scopes

- Text Generation
  - Creative Writing
  - Translation
  - Customer Support
- Code Generation
  - Code Generation
  - Code Compilation
- Visual Content:
  - Image Enhancement
  - Video Prediction
- Audio Generation
  - Music Composing
  - Text-to-Speech (TTS) Generation

Generative AI Tools

Some Popular Tools



Chat-GPT

DALLE / DALLE 2

Google Bard AI

Midjourney AI

Recurrent Neural Networks (RNNs)

Long Short-Term Memory (LSTM) Networks

Generative Adversarial Networks (GANs)

Variational Autoencoders (VAEs)

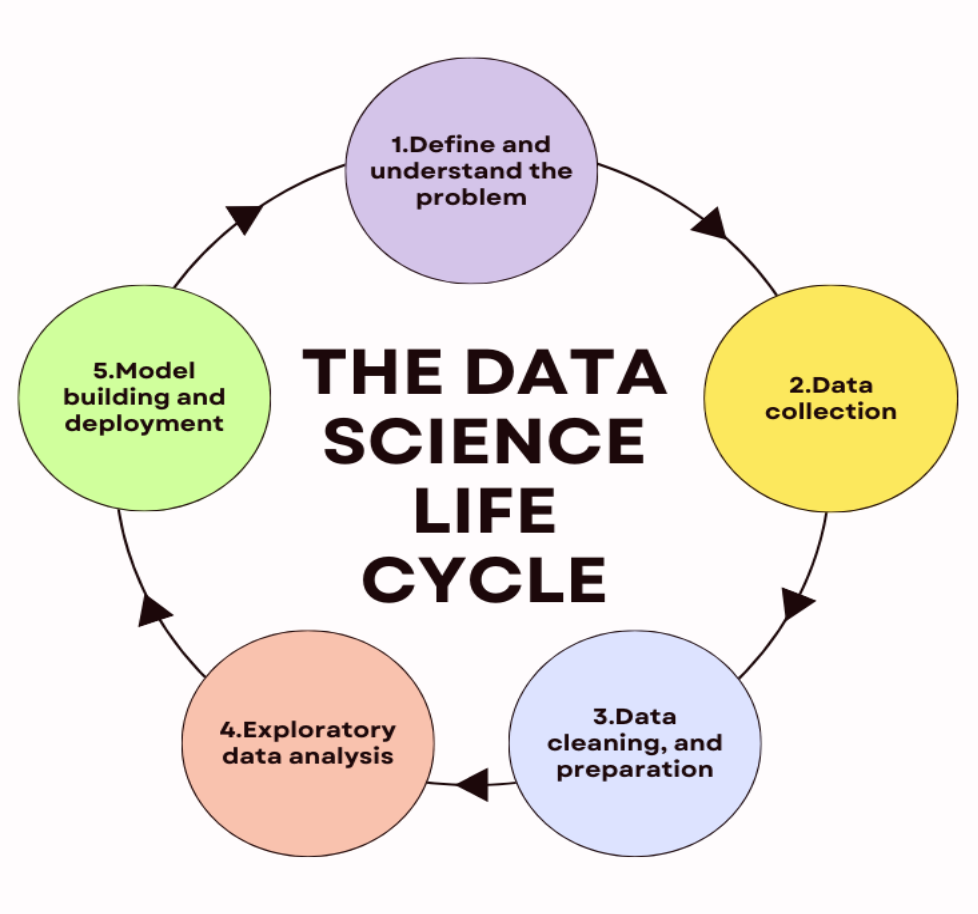
Transformer Networks

N-Gram Model

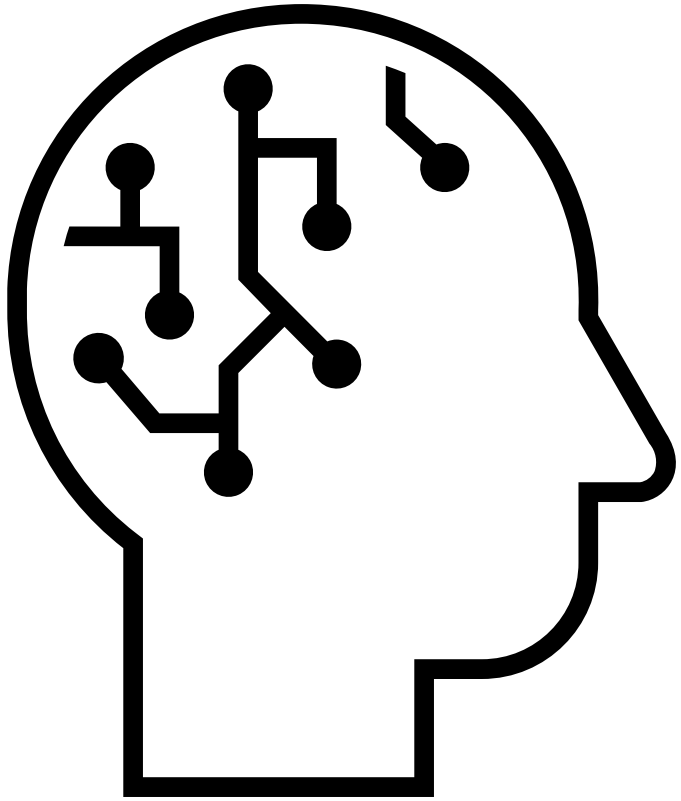
How to Use Data for a Smart Career?

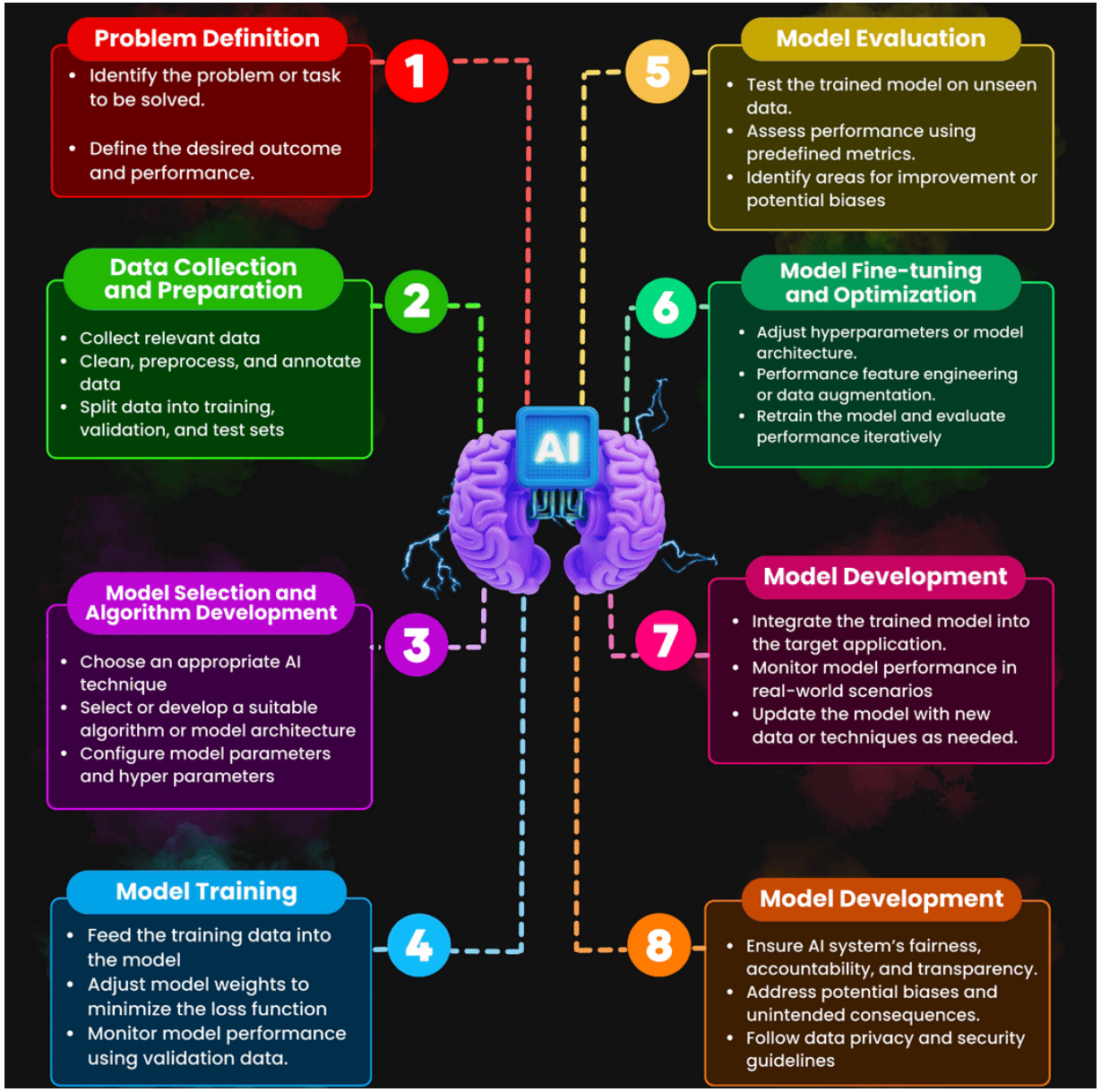
How to Use Data for a Smart Career?

Data Science Lifecycle



How to Build an AI Application



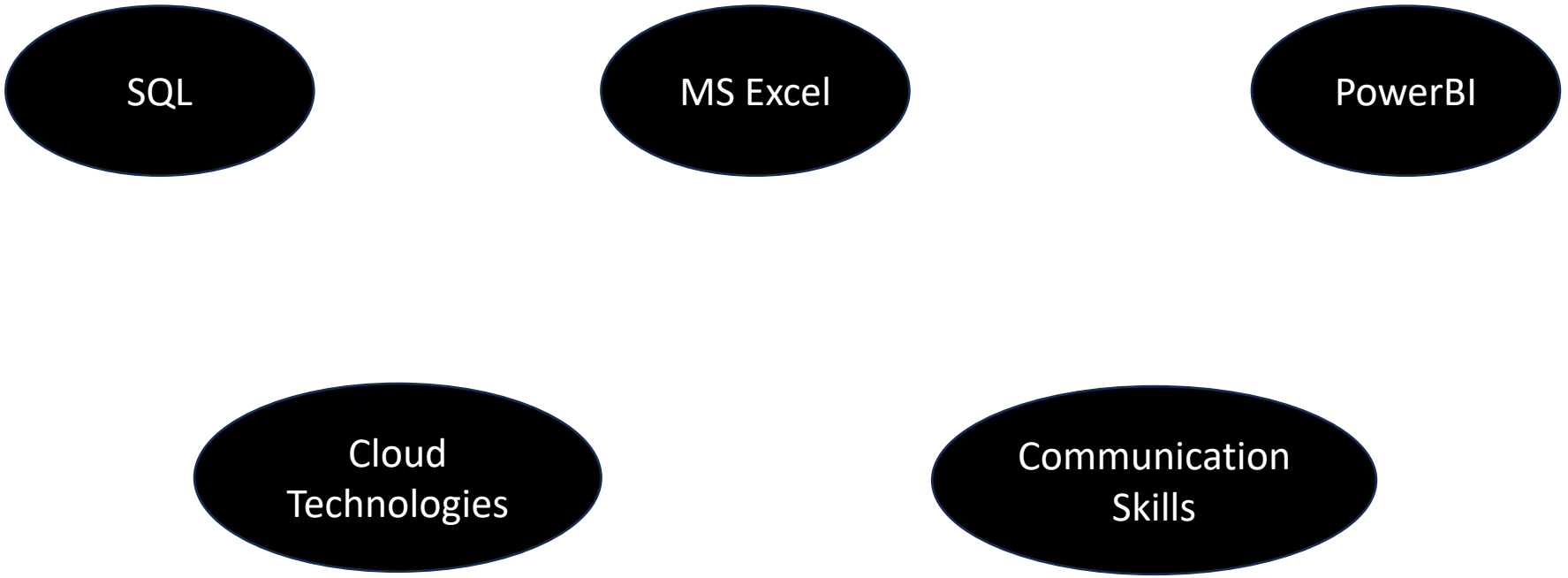


What Will We Learn?

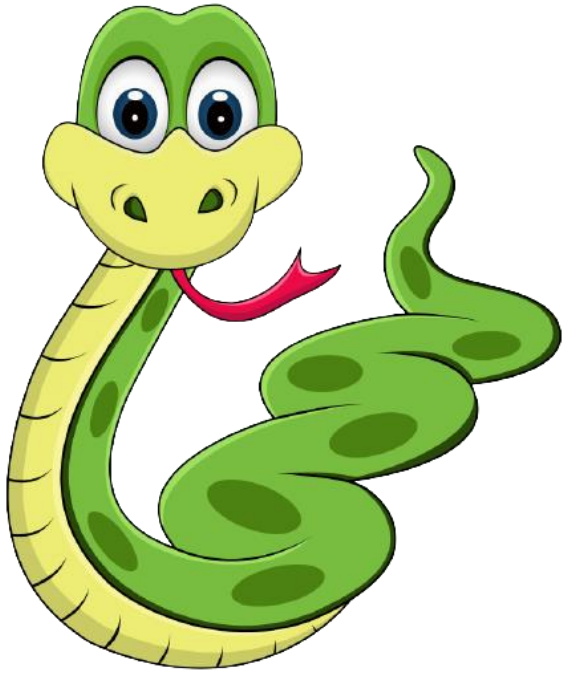
- Python
- Necessary Libraries
- Statistics for Data Science
- Linear Algebra for Data Science
- Partial Derivative in Maths
- Machine Learning Algorithms
  - Supervised
  - Unsupervised
- Deep Learning Algorithms
  - ANN
  - CNN
  - RNN
  - LSTM
- Algorithms Implementation
- Model Evaluation
- Model Optimization

Details Module

Some Common Skills not in This Course



Why Python for Data Science & AI



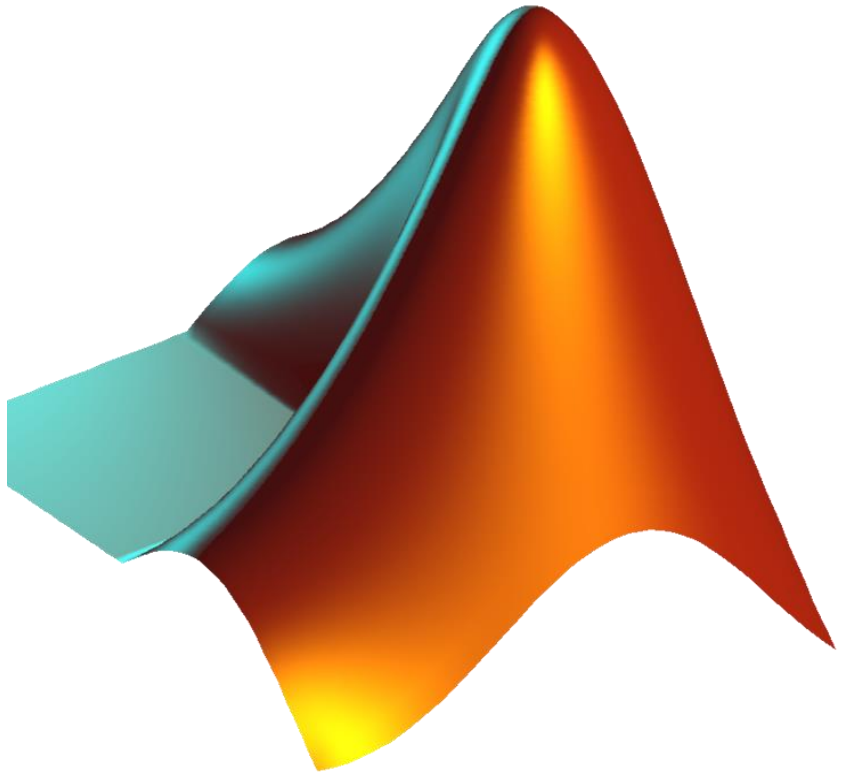
1. Easy Syntax, Flexible, Support OOP & Faster
2. Python has Machine Learning Libraries
3. Python has Data Analysis Library
4. Python has Data Frame Library
5. Python has Calculator Library
6. Python is Significant for Deep Learning
7. Keras, Tensorflow, Pytorch
8. Web (Django & Flask)
9. Open Resources

R for Data Science

- Data Analysis
- Data Visualization
- Statistical Libraries
- Statistics and Research Methods

MATLAB for AI

- Numerical Computing
- Matrix Operations
- Machine Learning Toolbox
- Deep Learning Support
- Data Visualization
- Simplicity and User-Friendly Interface
- Community Support
- Integration with Other Languages



Build Your Strong Social Network

[LinkedIn]

[GitHub]

kaggle

[Kaggle]

[Download the Books]

Thank you!